CS516: DATA WAREHOUSING AND DATA MINING (3-0-0: 3)

Introduction to Data Warehousing, Establishing Dimensional Modelling, Modelling the business process, Building fact tables Creating dimensions.

Building the Data Warehouse, Architecting the physical database, Ensuring data quality.
Extracting data from multiple sources.

Leveraging BI for Data Analysis, Deploying a Complete Data Warehouse Solution, Materialization techniques including full, partial and no materialization.

Basic concepts in data mining, Data measurement, exploratory data analysis, data visualization.

Predictive modelling: classification and regression, model fitting as optimization, evaluation of predictive performance, over-fitting, regularization, other data mining tasks: clustering and pattern detection, Lasso method for regularized regression. Information retrieval and search, Text classification, Unsupervised learning.

Recommender System: Recommender data, Netflix prize data, nearest neighbor algorithms, matrix decomposition algorithms, efficient algorithms for large data sets, modeling systematic effects.

Web Data Analysis: Web data: collection and interpretation, analyzing user browsing behavior, learning from click through data, predictive modeling and online advertising, link analysis and the Page-Rank algorithm. Approaches to Market Basket and quantitative data analysis using ARM (single and multi objective). Dynamic and distributed association mining.

Social Network Analysis: descriptive analysis of social networks, network embedding and latent space models, network data over time: dynamics and event-based networks, link prediction.

Text Books and References

1. D. J. Hand, H. Mannila, P. Smyth, “Principles of Data Mining”, MIT Press.
2. C.D. Manning, P. Raghavan, and H. Schutze, “Introduction to Information Retrieval”, Cambridge University Press.
3. R. Kimbal, “The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling”, Wiley.